



HAL
open science

ADL-dependent older adults were identified in medico-administrative databases

Emilie Hucteau, Pernelle Noize, Antoine Pariente, Catherine Helmer, Karine
Peres

► **To cite this version:**

Emilie Hucteau, Pernelle Noize, Antoine Pariente, Catherine Helmer, Karine Peres. ADL-dependent older adults were identified in medico-administrative databases. *Journal of Clinical Epidemiology*, 2021, 10.1016/j.jclinepi.2021.06.014 . hal-03322689

HAL Id: hal-03322689

<https://hal.science/hal-03322689>

Submitted on 5 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1 ADL-dependent older adults were identified in medico-administrative databases

2

3 Emilie Hucteau^{a,b}, Pernelle Noize^{b,c,d}, Antoine Pariente^{b,c,d}, Catherine Helmer^{a,b*}, Karine
4 Pérès^{b,e*}

5 * These two authors contributed equally to this work

6 ^a Univ. Bordeaux, Inserm UMR 1219, Bordeaux Population Health Research Center, team
7 Lifelong Exposure, Health and Aging, Bordeaux, France

8 ^b DRUGS-SAFE National Platform of Pharmacoepidemiology, Bordeaux, France

9 ^c Univ. Bordeaux, Inserm UMR 1219, Bordeaux Population Health Research Center, team
10 Pharmacoepidemiology, Bordeaux, France

11 ^d Bordeaux University Hospital, Public Health department, Medical pharmacoepidemiology,
12 Bordeaux, France

13 ^e Univ. Bordeaux, Inserm UMR 1219, Bordeaux Population Health Research Center, team
14 Psychoepidemiology of aging and chronic diseases, France

15

16 **E-mail addresses**

17 Emilie Hucteau: emilie.hucteau@u-bordeaux.fr

18 Pernelle Noize: pernelle.noize@u-bordeaux.fr

19 Antoine Pariente: antoine.pariante@u-bordeaux.fr

20 Catherine Helmer: catherine.helmer@u-bordeaux.fr

21 Karine Pérès: karine.peres@u-bordeaux.fr

22

23 **Corresponding author:**

24 INSERM U1219 Bordeaux Population Health Research Center
25 University of Bordeaux, 146, rue Léo Saignat,
26 F-33076 Bordeaux Cedex, France.
27 emilie.hucteau@u-bordeaux.fr

28

29

30

31

1 **ABSTRACT**

2

3 **Objective.** We aimed to develop an algorithm for the identification of basic Activities of Daily Living
4 (ADL)-dependency in health insurance databases.

5 **Study Design and Setting.** We used the AMI (Aging Multidisciplinary Investigation) population-based
6 cohort including both individual face-to-face assessment of ADL-dependency and merged health
7 insurance data. The health insurance factors associated with ADL-dependency were identified using a
8 LASSO logistic regression model in 1000 bootstrap samples. An external validation on a 1/97th
9 representative sample of the French Health Insurance general population of Affiliates has been
10 performed.

11 **Results.** Among 995 participants of the AMI cohort aged $\geq 65y$, 114 (11.5%) were ADL-dependent
12 according to neuropsychologists individual assessments. The final algorithm developed included: age,
13 sex, four drug classes (dopaminergic antiparkinson drugs, antidepressants, antidiabetic agents, lipid
14 modifying agents), three type of medical devices (medical bed, patient lifter, incontinence equipment),
15 four medical acts (GP's consultations at home, daily and non-daily nursing at home, transport by
16 ambulance) and four long-term diseases (stroke, heart failure, coronary heart disease, Alzheimer and
17 other dementia). Applying this algorithm, the estimated prevalence of ADL-dependency was 12.3% in
18 AMI and 9.5% in the validation sample.

19 **Conclusion.** This study proposes a useful algorithm to identify ADL-dependency in the health
20 insurance data.

21

22 **Keywords**

23 Dependency; Activities of daily living; Aged; Pharmacoepidemiology; Health insurance data; Cohort
24 study

25 **Running Title**

26 Identification of ADL-dependency in medico-administrative databases

27 **Word count**

28 3285 words

29

What is new?

Key findings

- The algorithm allows the identification of activities of daily living (ADL)-dependency among older adults in the health insurance databases.
- It includes drugs classes, medical devices and acts, long-term diseases variables and age and sex.
- Applying it to the French health insurance database as an external validation, it led to a prevalence of ADL-dependency of 9.5%, close to what is expected.

What this study adds?

- The algorithm was developed using a population-based cohort combining face-to-face assessment of ADL-dependency status and health insurance data.
- The findings confirmed the relevancy of this algorithmic approach to identify ADL-dependency in the French health insurance databases where this important status does not exist so far.

What is the implication and what should change now?

In the context of rapid aging of the population, the easy identification of ADL-dependency from health insurance data has important public health implications: to estimate its prevalence, its evolution over time and its determinants.

1. INTRODUCTION

The worldwide intensification of population aging led age-related diseases and their consequences to become a major public health concern [1]. According to the United Nations Highlights on World Population Ageing 2017, the over-80s will more than triple until 2050, then representing around 425 million people [1]. This will have major consequences in terms of healthcare and social care needs, relating to both the burden of the multiple chronic conditions it implies, and to their consequences in terms of dependency. Dependency is defined as a need for human assistance in activities of daily living (ADL) [2] and this is obviously the most severe stage of dependency that particularly mobilise the health, social and informal resources.

Whilst this dimension of health appears very important to monitor when considering population of older adults, most of the existing medico-administrative databases that are now routinely used for the assessment of health do not capture directly such information. Indeed, the detection of dependency ideally requires individual assessments performed by specially trained interviewers using standardised procedures and validated scales. If such expensive procedures have been deployed in the context of dedicated cohort studies, it cannot be implemented at the general population level. Despite massive amount of information, medico-administrative databases poorly contribute nowadays to the study of dependency. Yet, dependency is a relevant indirect indicator of health and its assessment in quasi-exhaustive samples of the older population would be crucial for public health to estimate the current and future needs of social and health care, which are expected to dramatically increase in the future. Such assessment could also allow to screen large populations of older adults in order to identify dependent people who could benefit from specific interventions. Finally, the identification of dependency in large unbiased databases would have applications for research purposes, for example to describe the path of dependent people, study the determinants of dependency, or controlling other analyses in administrative healthcare claims using dependency as a proxy of global health. The detection of dependency has previously been performed through the developing of dedicated algorithms from US claims data, but in a population and health system limiting their applicability to the very different European setting [3–5].

The objective of this study was to develop an algorithm allowing to identify older ADL-dependent adults in French Health Insurance databases using population-based cohort combining individual assessment of ADL-dependency and health insurance data.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30

2. METHODS

2.1. Study population and data collection

The AMI (Aging Multidisciplinary Investigation) study is a prospective population-based cohort on health and aging conducted in former farmers living in rural environment in Gironde department (South-Western France) [6]. A total of 1002 individuals aged 65 or older, retired from agriculture and affiliated to the French Farmer Health Insurance System (“Mutualité Sociale Agricole”, MSA) were included in 2007-2009. At baseline (T0), trained neuropsychologists used standardized questionnaires at the participant's home to collect data (socio-demographics characteristics, living conditions, lifestyle, health status, drugs use) and to assess dependency. Participants were interviewed in the same conditions two years later in 2010-2011 (T2). In addition to these data, the MSA provided health insurance data for the cohort participants.

2.2. Definition and assessment of ADL-dependency

Among the different domains of activity limitation, we chose to focus on the most severe one, the basic activities of daily living (ADL). The neuropsychologists used the Katz's scale [7] to assess participants' ADL-dependency level during the home interviews. This scale includes five different activities: dressing, bathing, toileting, transferring and eating. A participant was considered as dependent if he/she could not perform at least one activity without a given level of human assistance (personal assistance, directive assistance or supervision). This assessment was used as gold standard. To increase the power of our analysis, we considered dependency either at baseline (T0) or 2 years later (T2) (Supplementary Figure 1). Accordingly, in the analysis, we used: i) T0 data for participants who were independent at T0 and remained independent at T2 and for those who were already dependent at T0; ii) and T2 data for participants with incident dependency at T2 and for those who were dependent at T2 with missing information on dependency at T0. Date of the assessment was considered as the inclusion date for this study.

2.3. Health-related reimbursements from the merged French Farmer Health Insurance System

For each participant, all health-related reimbursements were extracted from the French Health Insurance System database and linked to the cohort data. For this analysis we considered health-related reimbursements that were concomitant with the dependency status. The information provided included: out-hospital medications data (all reimbursements performed), out-hospital medical and

1 paramedical acts, medical equipment and devices and the registered Long-Term Diseases (LTD) (30
2 specific LTDs allowing 100% health expenditures coverage by the Insurance System, supplementary
3 Table 1). The data included in the analysis were selected based on expert's opinion (co-authors of this
4 paper: experts in epidemiology of aging and dependency (CH, KP) and in pharmacoepidemiology with
5 a high experience of health insurance databases (PN, AP)) and the exposure period considered for
6 each factor was defined according to the type of the data (supplementary Table 2). The LTD being
7 mainly irreversible, we considered that a person identified as suffering from a LTD at a time remained
8 diseased for the rest of the follow-up.

9 Drugs and LTD that concerned at least 20 participants and medical acts and devices that concerned
10 at least 5 participants were included in our analyses. After adding age (< 75 / $75- < 85$ / ≥ 85) and sex,
11 70 factors were thus included in the analysis for the algorithm development: 39 drugs, 10 medical and
12 paramedical acts, 9 medical equipment and devices and 10 LTD (Figure 1).

13 **2.4. Data analysis**

14 *2.4.1. Creation of the algorithm of ADL-dependency*

15 Participants' main characteristics (sociodemographic, drugs consumption, medical and paramedical
16 acts, medical equipment and devices and LTD) were described according to the ADL-dependency
17 status using bivariate analyses with Pearson's chi-square and Fisher exact test.

18 To develop the algorithm, we first identified the factors associated with ADL-dependency. Because of
19 a high number of factors, we used a logistic regression model with LASSO regularization repeated for
20 1000 bootstrap samples [8–10]. Over these 1000 iterations, the frequency of selection was recorded
21 for each factor. Controlling for age and sex, three provisional algorithms were built as the sum of each
22 variable multiplied by its beta coefficient from the variables selected in 90% (algorithm90), 80%
23 (algorithm80), and 70% (algorithm70) of the bootstrap samples and different cut-offs were considered
24 for each algorithm. Using estimated prevalence, the area under the receiver operating characteristics
25 (ROC) curve (AUC), sensitivity (Se), specificity (Sp), positive and negative predictive value (PPV and
26 NPV respectively), and Youden index, we identified the most accurate algorithm out of the three; for
27 this, we then determined the most accurate cut-off for ADL-dependency identification.

28 *2.4.2. Application and external validation on the French General Health Insurance database*

29 The "Echantillon Généraliste de Bénéficiaires (EGB)" is an anonymous 1/97th sample of the French
30 health insured population [11]. Established in 2005, it is representative of the general population in

1 terms of sex and age and contains for each affiliate an exhaustive recording of all out-hospital health
2 expenditures (including drug deliveries, medical and paramedical acts, and reimbursements for
3 medical equipment and devices), hospitalisation data, and information on LTD.

4 For the validation phase, we selected persons aged 65 and over and affiliated to the Health Insurance
5 General Scheme (covering around 85% of the population) in January 2010 (n=74,652, Supplementary
6 Figure 2). In this sample, we applied the developed algorithm quarterly from 2010 to 2015. Thus, the
7 same population being followed over the 6 years, we obtained for a person followed until December
8 2015, 24 dependency status, one by quarter. The end of the follow-up was defined as the date of
9 death, the date of end of the affiliation, or December 2015 the 31st, whichever came first. Yet, applying
10 the algorithm by quarter may lead to an unstable status due to a transitory increased or decreased
11 consumption of care. However, the level being severe (with the Katz's scale) [12] and ADL-
12 dependency generally consisting in stable dependency, we assumed its reversibility to be very low.

13 We thus established a strategy for the use of the algorithm in the database considering both the
14 potentiality of an unstable status and the limitations of the data. First, we applied several rules to
15 smooth the results: i) if, while considered dependent over the preceding and following quarters, a
16 person was non-dependent in one quarter owing to the algorithm results, the person's status was
17 changed for this quarter into "dependent"; ii) similarly, if a person was considered dependent for a
18 quarter but not the two preceding and following ones, the person's status for this quarter was changed
19 for "non-dependent" (Figure 2). Second, as the reimbursement data (drugs, medical acts,
20 equipment...) are not available during a period of hospitalization, if a person was dependent the
21 quarter before a long period of hospitalization (30 days or more), we considered that the person
22 remained dependent during this hospitalization. Third, we classified some people as dependent
23 regardless of the algorithm: 1) people who were admitted to hospital from long-term care or from a
24 medico-social housing structure (including nursing homes for which information on care received is
25 lacking from the database) or who left hospital towards these structures; and 2) people hospitalized for
26 at least 3 months. For these people, the beginning of the dependency period was the quarter where
27 the hospitalization started.

28 Finally, to confirm the relevance of the algorithm: i) we estimated the prevalence of dependency
29 according to age and sex to check for consistency with literature; ii) we evaluated the reversibility rates
30 to ensure that this rate was low as expected for ADL-dependency; and iii) we calculated death rate at

1 the end of the follow-up period according to the dependency status (among dependent in 2010 vs non-
2 dependent in 2010) and compared it to AMI death rate at the 6-year follow-up (among dependent at
3 T0 or T2 vs non-dependent at these times).

4 Analyses were performed using SAS 9.4 (*SAS Institute, Inc., Cary NC*) and R 3.4.1.

5 **3. RESULTS**

6 ***3.1. Description of the AMI cohort***

7 Among the 1002 participants, 995 were included in our analyses after exclusion of 7 individuals with
8 missing data on ADL scale: 75 were dependent at T0, 39 were incident dependent at T2 and 881 were
9 independent at T0 and T2, leading to 11.5% of dependent participants in our sample. Mean age of
10 participants was 76.4 (standard deviation: 6.7) and 37.5% were female. Main characteristics of the
11 participants are shown in Table 1 according to the ADL-dependency status.

12 ***3.2. Development of the algorithm***

13 Aside of age and sex that were systematically considered, among the 70 variables initially included, 19
14 were selected in at least 70% of the bootstrap samples, 15 in at least 80% and 10 in at least 90%
15 (Figure 3). For the three provisional algorithms created, the AUC was excellent: 0.92 for algorithm90
16 and 0.94 for algorithm70 and algorithm80. Different cut-offs were tested for each algorithm; the most
17 performant ones are shown in Table 2. Although algorithm70 at 2.5 cut-off had a better Youden Index,
18 its VPP was quite small and its estimated prevalence quite high compared to the expected prevalence.
19 Thus, the retained algorithm, showing the best compromise between estimated prevalence and
20 performances, was algorithm80 at a 3.1 cut-off with a Se at 75.4%, a Sp at 95.9%, a PPV at 70.5%
21 and a NPV at 96.8%. The average of this algorithm was 0.8 (SD=1.1) for non-dependent and 5.7
22 (SD=3.6) for dependent persons. The final algorithm corresponding to algorithm80 included 17
23 variables (Figure 2).

24 The prevalence of ADL-dependency based on the algorithm with the defined 3.1 cut-off was estimated
25 at 12.3% from the French Farmer Health Insurance System merged to the AMI cohort vs. 11.5%
26 based on the individual ADL assessment of cohort participants.

27 ***3.3. External validation on the EGB sample***

28 In the EGB, 74,652 persons met the inclusion criteria (Supplementary Figure 2). Among them, 59.0%
29 were female, and mean age was 76.3 (SD=7.7). The average follow-up was 5.3 years (SD=1.4) for a
30 maximum of 6 years.

1 Among the 74,652 persons, 9.5% were classified as dependent during 2010 first quarter. This
2 prevalence increased with age: 2.4% among people aged less than 75, 10.4% among those aged 75-
3 84 and 30.7% among those aged 85 or more. With the aging of the studied sample between 2010
4 (first quarter) and 2015 (last quarter) (same population followed-up over the 6 years), the prevalence
5 of dependency increased as expected to reach 16.3%. Over the 6-year follow-up period, 27.9% were
6 considered ADL-dependent at least once, with a higher frequency in females (31.5% vs 22.6% in
7 males). Among persons classified as dependent at least once, 80.0% remained dependent after the
8 first dependency period; this stable status was higher in males than in females (82.7% and 78.7%
9 respectively). On average, men became dependent at younger ages than women (81.6 years
10 (SD=6.8) vs. 84.0 years (SD=6.7) respectively) probably due to a lower life expectancy of about 6
11 years in men than in women in France [13]. Finally, death rate was 73.7% among dependent EGB
12 people (vs 22.7% among non-dependent) and 77.2% among dependent AMI participants (vs 22.1%
13 among non-dependent).

14 **4. DISCUSSION**

15 Using a population-based cohort combining individual face-to-face assessment of ADL-dependency
16 (as gold standard) and health insurance data, we developed an algorithm to identify ADL-dependency
17 in administrative healthcare claims. In addition to age and sex, this algorithm retained 15 variables
18 related to health care (drugs, medical and paramedical acts and medical devices) and LTD. In the AMI
19 cohort, the developed algorithm presented excellent performances, with an AUC of 0.94; the
20 identification of ADL-dependency using the developed algorithm allowed obtaining a prevalence
21 similar to that observed in participants (12.3% vs. 11.5%). Within the external validation in a large
22 sample of the French older adults insured population aged 65 and over, the obtained prevalence was
23 lower, of 9.5%.

24 In the context of intense aging of the populations [1], this algorithm, easy to apply in administrative
25 healthcare claims, may be valuable in terms of public health and population health monitoring,
26 allowing to provide key indicators of prevalence, incidence, transition probabilities, time spent in
27 dependency, or DALYs. It also offers excellent opportunity to estimate the costs of dependency in
28 medico-economic analyses, to follow trends over time and to perform projections; crucial data to
29 anticipate and organise the society to face future health care and social care needs associated to the
30 intensive aging of the population. Moreover, it could also allow studying the determinants of

1 dependency, including health events and drug consumption, and identifying individuals at-risk of
2 dependency who could benefit from targeted preventive strategies.

3 In the dependency process, different levels of increasing severity exist according to the domains of
4 activity limitation hierarchically affected (first mobility, then instrumental ADL and finally basic ADL) [2,
5 12]. For the development of this algorithm, we chose to focus exclusively on basic activities for several
6 reasons. First, this is the level of dependency associated to the highest levels of medical, social,
7 informal care, caregiver's burden [13, 14] and costs [15, 16] and poorer quality of life of the elderly
8 [17]. Second, we assumed that less severe stages of dependency were less likely to be identified from
9 health expenditures.

10 To our knowledge, all the algorithms existing to date for the identification or prediction of dependency
11 have been developed in the United-States using Medicare [3–5]. Davidoff et al. developed and
12 validated an algorithm aiming to detect a summary measure of dependency considering globally all
13 stages of dependency (IADL and ADL) [3]. Combining different levels of dependency in a single
14 indicator leads to a heterogeneous group with different associated risks, costs and needs. As in the
15 present work, Faurot et al. developed an algorithm to detect basic ADL-dependency among older
16 adults, used as a proxy of frailty in order to improve unmeasured confounding in the analyses
17 conducted using healthcare claims [4]; this algorithm has recently been validated within the ARIC
18 population-based study [18]. Finally, Kinoshian et al. recently tested an already existing commercialized
19 algorithm (whose development has not been published), to evaluate its performances for ADL
20 dependency detection within the National Long-Term Care Survey [5]. With an AUC at 0.94 in the
21 developing sample, our algorithm slightly outperforms these previous algorithms (AUC of 0.80 to 0.92
22 in the estimation samples, and of 0.71 to 0.92 in the validation samples) [3, 5, 18]. Additionally, the
23 use of LASSO logistic regression model allowed us obtaining a lighter algorithm including only 17
24 variables, thus theoretically easier to apply in health insurance databases.

25 Contrary to the other algorithms, we chose to include the drugs used, instead of only including
26 diagnoses, similarly to what was done in adaptations of comorbidity score to health insurance
27 databases [19, 20]. The LASSO logistic regression model allowed us to include all the drug classes in
28 the model, without having any a priori hypothesis on drugs that could improve the detection of
29 dependency. At the end, four drug classes were retained in the final algorithm. Several of the variables
30 identified in previous algorithms, such as nursing visit, home visit, ambulance, hospital bed, dementia,

1 stroke, heart failure were similarly retained in our algorithm. In addition, we also retained dopaminergic
2 antiparkinson drugs, antidiabetic agents and lipid modifying agents, as well as incontinence
3 equipment, whereas Faurot et al. retained diagnosis of Parkinson's disease, diabetic complications,
4 lipid abnormalities, and bladder continence [4]. Contrary to other algorithms [3, 4], wheelchair was not
5 retained in the final algorithm, even it would have been if we had selected algorithm70. However, our
6 statistical approach probably led to retain other variables in the model, highly correlated with
7 wheelchair use.

8 The main strength of our study is its design relying on the combined use in the same study of
9 information on individual assessment of dependency performed by trained neuropsychologists using a
10 validated scale and information on health care obtained from merged health insurance data
11 exhaustive in terms of outpatient care and hospitalisation episodes. Given the major present and
12 future implications of ALD-dependency, this study provides an indicator particularly relevant in a public
13 health perspective. However our study also has some limits. The algorithm has been developed on a
14 specific older population of retired farmers, which explains for instance the specific sex-ratio in favor to
15 men. Medical acts and prescriptions may thus differ from other populations. Also, the sample used to
16 develop the algorithm only included 995 participants, of which 114 were ADL-dependent. However,
17 this did not preclude developing an algorithm with excellent AUC, which application on an external
18 sample allowed obtaining results for prevalence and age at ADL-dependency entry consistent with the
19 literature [13, 22, 23]. Moreover, as a validation step of our algorithm, we found much higher mortality
20 rate for those identified as dependent compared to those nondependent in the EGB, and this higher
21 rate was very close to that found in the AMI cohort among dependent participants. This algorithm
22 based on French health prescriptions and patients medical and paramedical paths, may be not directly
23 transposable to other countries, all having specific health and social care insurances and procedures.
24 The same would be true for the strategy of use we developed. This was constrained by the lack of in-
25 hospital or nursing home care data that might be available in health insurance databases in other
26 countries. The classification rules we retained could be adapted and improved in such settings.

27 **5. CONCLUSION**

28 Our algorithm allowing the identification of ADL-dependency from health insurance databases might
29 constitute a valuable tool for public health and population health monitoring from electronic healthcare
30 databases. The easy to obtain information its use could provide regarding dependency prevalence,

- 1 incidence, or determinant identification represents key data for policy makers to anticipate and adapt
- 2 societies to aging and if possible to develop preventive strategies.
- 3

ACKNOWLEDGMENTS

1 This study is part of the DRUGS-SAFE research program, funded by the French Medicines Agency
2 (*Agence Nationale de Sécurité du Médicament et des Produits de Santé*, ANSM). This publication
3 represents the views of the authors and does not necessarily represent the opinion of the French
4 Medicines Agency. The sponsor has no involvement in any parts of this work.

5 The AMI project was supported by the Association de Gestion pour le Compte des Institutions
6 Complémentaires Agricoles, the Caisse Mutuelle Autonome de Retraités Complémentaires Agricoles,
7 the Caisse de Retraite Complémentaire des Cadres de l'Agriculture, the Caisse Centrale de
8 Prévoyance Mutuelle Agricole Prévoyance, the Caisse de Prévoyance des Cadres d'Entreprises
9 Agricoles, Agri Prévoyance, the Mutualité Sociale Agricole de Gironde, the Caisse Centrale de la
10 Mutualité Sociale Agricole, and the Caisse Nationale de Solidarité pour l'Autonomie.

11 **Declaration of interest:** none

12

13 **Author CRediT Statement:** Emilie Hucteau: Methodology, formal analysis, software, writing – original
14 draft, writing – review & editing. Pernelle Noize: methodology, data interpretation, writing – review &
15 editing. Antoine Pariente: data interpretation, writing – review & editing. Catherine Helmer:
16 conceptualization, methodology, data interpretation, writing – review & editing. Karine Pérès:
17 conceptualization, investigation, methodology, data interpretation, writing – review & editing.

1 REFERENCES

- 2 [1] United Nations, Department of Economic and Social Affairs, Population Division, "World
3 population ageing 2017," 2017. [Online]. Available:
4 [https://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2017_Report.p](https://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2017_Report.pdf)
5 [df](https://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2017_Report.pdf).
- 6 [2] A. Edjolo, C. Proust-Lima, F. Delva, J.-F. Dartigues, and K. Pérès, "Natural History of
7 Dependency in the Elderly: A 24-Year Population-Based Study Using a Longitudinal Item
8 Response Theory Model," *Am. J. Epidemiol.*, vol. 183, no. 4, pp. 277–285, Feb. 2016, doi:
9 10.1093/aje/kwv223.
- 10 [3] A. J. Davidoff *et al.*, "A novel approach to improve health status measurement in observational
11 claims-based studies of cancer treatment and outcomes," *J. Geriatr. Oncol.*, vol. 4, no. 2, pp.
12 157–165, Apr. 2013, doi: 10.1016/j.jgo.2012.12.005.
- 13 [4] K. R. Faurot *et al.*, "Using Claims Data to Predict Dependency in Activities of Daily Living as a
14 Proxy for Frailty," *Pharmacoepidemiol. Drug Saf.*, vol. 24, no. 1, pp. 59–66, Jan. 2015, doi:
15 10.1002/pds.3719.
- 16 [5] B. Kinoshian, D. Wieland, X. Gu, E. Stallard, C. S. Phibbs, and O. Intrator, "Validation of the JEN
17 frailty index in the National Long-Term Care Survey community population: identifying
18 functionally impaired older adults from claims data," *BMC Health Serv. Res.*, vol. 18, no. 1, p.
19 908, Nov. 2018, doi: 10.1186/s12913-018-3689-2.
- 20 [6] K. Pérès *et al.*, "Health and aging in elderly farmers: the AML cohort," *BMC Public Health*, vol. 12,
21 p. 558, Jul. 2012, doi: 10.1186/1471-2458-12-558.
- 22 [7] S. Katz, T. D. Downs, H. R. Cash, and R. C. Grotz, "Progress in development of the index of
23 ADL," *The Gerontologist*, vol. 10, no. 1, pp. 20–30, 1970, doi: 10.1093/geront/10.1_part_1.20.
- 24 [8] M. Pavlou, G. Ambler, S. Seaman, M. De Iorio, and R. Z. Omar, "Review and evaluation of
25 penalised regression methods for risk prediction in low-dimensional data with few events," *Stat.*
26 *Med.*, vol. 35, no. 7, pp. 1159–1177, Mar. 2016, doi: 10.1002/sim.6782.
- 27 [9] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *J. R. Stat. Soc. Ser. B*
28 *Methodol.*, vol. 58, no. 1, pp. 267–288, 1996, doi: 10.1111/j.2517-6161.1996.tb02080.x.
- 29 [10] P. Guo and Y. Hao, *SparseLearner: Sparse Learning Algorithms Using a LASSO-Type Penalty*
30 *for Coefficient Estimation and Model Prediction*. 2015.
- 31 [11] P. Tuppin, L. de Roquefeuil, A. Weill, P. Ricordeau, and Y. Merlière, "French national health
32 insurance information system and the permanent beneficiaries sample," *Rev. D'Épidémiologie*
33 *Santé Publique*, vol. 58, no. 4, pp. 286–290, Aug. 2010, doi: 10.1016/j.respe.2010.04.005.
- 34 [12] P. Barberger-Gateau, C. Rainville, L. Letenneur, and J. F. Dartigues, "A hierarchical model of
35 domains of disablement in the elderly: a longitudinal approach," *Disabil. Rehabil.*, vol. 22, no. 7,
36 pp. 308–317, May 2000, doi: 10.1080/096382800296665.
- 37 [13] K. Pérès, A. Edjolo, J.-F. Dartigues, and P. Barberger-Gateau, "Recent Trends in Disability-Free
38 Life Expectancy in the French Elderly. Twenty Years Follow-Up of the Paquid Cohort," *Annu.*
39 *Rev. Gerontol. Geriatr.*, vol. 33, pp. 293-311(19), Feb. 2013, doi: 10.1891/0198-8794.33.293.
- 40 [14] N. B. Coe, M. M. Skira, and E. B. Larson, "A Comprehensive Measure of the Costs of Caring for
41 a Parent: Differences According to Functional Status," *J. Am. Geriatr. Soc.*, vol. 66, no. 10, pp.
42 2003–2008, 2018, doi: 10.1111/jgs.15552.
- 43 [15] S. Andrieu *et al.*, "[Burden experienced by informal caregivers assisting Alzheimer's patients in
44 the REAL.FR study]," *Rev. Med. Interne*, vol. 24 Suppl 3, pp. 351s–359s, Oct. 2003, doi:
45 10.1016/s0248-8663(03)80695-1.
- 46 [16] L. I. van Lier *et al.*, "Predictors of Societal Costs of Older Care-Dependent Adults Living in the
47 Community in 11 European Countries," *Health Serv. Insights*, vol. 12, p. 1178632918820947,
48 2019, doi: 10.1177/1178632918820947.
- 49 [17] C. Lindholm, A. Gustavsson, L. Jönsson, and A. Wimo, "Costs explained by function rather than
50 diagnosis--results from the SNAC Nordanstig elderly cohort in Sweden," *Int. J. Geriatr.*
51 *Psychiatry*, vol. 28, no. 5, pp. 454–462, May 2013, doi: 10.1002/gps.3844.
- 52 [18] L. Christiansen, J. Sanmartin Berglund, C. Lindberg, P. Anderberg, and L. Skär, "Health-related
53 quality of life and related factors among a sample of older people with cognitive impairment,"
54 *Nurs. Open*, vol. 6, no. 3, pp. 849–859, Mar. 2019, doi: 10.1002/nop2.265.
- 55 [19] C. C. Cuthbertson *et al.*, "Controlling for Frailty in Pharmacoepidemiologic Studies of Older
56 Adults: Validation of an Existing Medicare Claims-based Algorithm," *Epidemiol. Camb. Mass*,
57 vol. 29, no. 4, pp. 556–561, 2018, doi: 10.1097/EDE.0000000000000833.
- 58 [20] A. Bannay *et al.*, "The Best Use of the Charlson Comorbidity Index With Electronic Health Care
59 Database to Predict Mortality," *Med. Care*, vol. 54, no. 2, pp. 188–194, Feb. 2016, doi:
60 10.1097/MLR.0000000000000471.

- 1 [21] S.-G. Béland, C. Tannenbaum, T. Ducruet, M. Prévaille, D. Berbiche, and Y. Moride, "Effect of
2 external variables on the performance of the geriatric comorbidity score derived from prescription
3 claims in the community-dwelling elderly," *Drugs Aging*, vol. 29, no. 11, pp. 891–897, Nov. 2012,
4 doi: 10.1007/s40266-012-0022-3.
- 5 [22] R. H. Harwood, A. A. Sayer, and M. Hirschfeld, "Current and future worldwide prevalence of
6 dependency, its relationship to total population, and dependency ratios.," *Bull. World Health
7 Organ.*, vol. 82, no. 4, pp. 251–258, Apr. 2004.
- 8 [23] M. Duée and C. Rebillard, "La dépendance des personnes âgées : une projection en 2040,"
9 *Données Soc. - Société Fr.*, vol. 7, pp. 613–9, 2006.
- 10
11

1 TABLES AND FIGURES

2 **Figure 1.** Selection of factors included in the analysis

3 *The figure should be read as follows: among all medical and paramedical acts, 12 were selected by*
4 *experts' opinion and, after exclusion of factors that concerned less than 5 participants, 10 were*
5 *included in the LASSO analysis.*

6 ¹ *In the ATC (Anatomical Therapeutic Chemical) classification drugs are divided into groups according*
7 *to the organ or system on which they act or their therapeutic and chemical characteristics; drugs are*
8 *coding on seven characters (letters and digits). The first level (first letter) defines the anatomical group*
9 *among 14 different ones. The second level (first two digits) gives the main pharmacological or*
10 *therapeutic subgroup and the third level (second letter) corresponds to chemical, pharmacological or*
11 *therapeutic subgroups. In the analysis we used the second level but for drugs of the nervous system*
12 *where the third level was considered due to their potential high relation with dependency.*

13 **Figure 2.** Algorithm to identify quarters of ADL-dependency

14 **Figure 3.** Factors selected by LASSO bootstraps

15 **Figure 4.** Prevalence of ADL-dependency among the EGB patients in 2010 and 2015 first quarter and
16 2015 last quarter

17 **Table 1.** Characteristics of the participants according to their ADL-dependency status. AMI cohort
18 2007-2011, N=995

19 *Only variables significantly associated to ALD-dependency ($p < 0.05$) are presented*

20 **Table 2.** Performance of the scores selected in 90%, 80% and 70% of the bootstrap samples
21 according to different cut-offs

22 *Se: sensitivity; Sp: specificity; PPV: Positive predictive value; NPV: Negative predictive value*

23 *Youden index is calculated as follows: $(Se+Sp)-1$*

24 *Algorithms at 90%, 80% and 70% refer to the factors selected in 90%, 80% and 70% of the bootstrap*
25 *samples for their construction*

26
27 **Supplementary Figure 1.** Selection of AMI participants included in the analysis

28
29 **Supplementary Figure 2.** Flow Chart for the EGB patients, an anonymous 1/97th sample of the
30 French health insured population

31 **Supplementary Table 1.** List of the thirty long-term diseases

32 **Supplementary Table 2.** Medical, paramedical acts and medical equipment and devices included in
33 the analysis with exposure period considered

34 *N: number of factors included*

35 *ATC: Anatomical, Therapeutic and Chemical) classification system*

36 *GP: General practitioner*

37 *For example, a person was considered exposed to "GP consultation" if he had at least 3 consultations*
38 *within the three-month period before the inclusion date.*

39

1 Table 1. Characteristics of the participants according to their ADL-dependency status. AMI cohort
 2 2007-2011, N=995.

Characteristics	Dependent n=114		Independent n=881	
	n (%)		n (%)	
Sociodemographic				
Sex				
Men	59	(51.8)	563	(63.9)
Women	55	(48.2)	318	(36.1)
Age				
less than 75	13	(11.4)	443	(50.3)
75 to less than 85	61	(53.5)	370	(42.0)
85 and over	40	(35.1)	68	(7.7)
Drugs consumption				
A02: Drugs for acid related disorders	43	(37.7)	191	(21.7)
A03: Drugs for functional gastrointestinal disorders	17	(14.9)	52	(5.9)
A06: Drugs for constipation	23	(20.2)	40	(4.5)
A10: Drugs used in diabetes	21	(18.4)	97	(11.0)
A12: Mineral supplements	13	(11.4)	34	(3.9)
B01: Antithrombotic agents	46	(40.4)	251	(28.5)
B03: Antianemic preparations	8	(7.0)	20	(2.3)
C03: Diuretics	40	(35.1)	161	(18.3)
C10: Lipid modifying agents	24	(21.0)	279	(31.7)
D01: Antifungals for dermatological use	11	(9.7)	11	(1.3)
H03: Thyroid therapy	12	(10.5)	49	(5.6)
M02: Topical products for joint and muscular pain	9	(7.9)	29	(3.3)
N02B: Other analgesics and antipyretics	34	(29.8)	163	(18.5)
N03A: Antiepileptics	12	(10.5)	30	(3.4)
N04B: Dopaminergic agents for Parkinson	13	(11.4)	9	(1.0)
N05A: Antipsychotics	14	(12.3)	11	(1.3)
N05B: Anxiolytics	22	(19.3)	65	(7.4)
N05C: Hypnotics and sedatives	19	(16.7)	58	(6.6)
N06A: Antidepressant drugs	31	(27.2)	52	(5.9)
N06D: Anti-dementia drugs	21	(18.4)	38	(4.3)
Medical and paramedical acts				
Daily nursing care	13	(11.4)	1	(0.1)
Non-daily nursing care	14	(12.3)	3	(0.3)
Home treatment accessories	25	(21.9)	5	(0.6)
Ambulance	7	(6.1)	2	(0.2)
Medical vehicle	5	(4.4)	12	(1.4)
At-home GP ¹ visits	52	(45.6)	71	(8.1)
Medical equipment and devices				
Walker	3	(2.6)	4	(0.4)
Anti-bedsore mattress	13	(11.4)	4	(0.4)
Wheelchair	12	(10.5)	1	(0.1)

Patient lifter	8	(7.0)	0	(0.0)
Medical bed	46	(40.3)	9	(1.0)
Nutrients	5	(4.4)	9	(1.0)
Drip equipment	4	(3.5)	2	(0.2)
Equipment for incontinence	15	(13.2)	7	(0.8)
<i>Long-term diseases</i>				
Stroke	9	(7.9)	17	(1.9)
Severe heart failure	25	(21.9)	90	(10.2)
Coronary disease	22	(19.3)	79	(9.0)
Alzheimer's disease and other dementia	20	(17.5)	7	(0.8)
Long-term psychiatric conditions	8	(7.0)	13	(1.5)

1 Only variables significantly associated to ADL-dependency ($p < 0.05$) are presented

1 Table 2. Performance of the algorithms selected in 90%, 80% and 70% of the bootstrap samples
 2 according to different cut-offs

Algorithm	Cut-off	Se (%)	Sp (%)	PPV (%)	NPV (%)	Youden (%)	Prevalence (%)
90%	2.5	75.4	93.9	61.4	96.7	69.3	14.1
	3.0	71.1	96.7	73.6	96.3	67.8	11.1
	3.5	64.9	97.4	76.3	95.5	62.3	9.7
80%	2.5	79.8	91.6	55.2	97.2	71.4	16.6
	3.0	75.4	95.8	69.9	96.8	71.2	12.4
	3.1	75.4	95.9	70.5	96.8	71.4	12.3
	3.2	73.7	96.1	71.2	96.6	69.8	11.9
	3.3	72.8	96.5	72.8	96.5	69.3	11.5
	3.4	70.2	96.5	72.1	96.2	66.7	11.2
	3.5	70.2	96.7	73.4	96.2	66.9	11.0
70%	2.5	78.1	95.2	67.9	97.1	73.3	13.2
	3.0	73.7	96.5	73.0	96.6	70.2	11.6
	3.5	69.3	98.0	81.4	96.1	67.3	9.7

3 Se: sensitivity; Sp: specificity; PPV: Positive predictive value; NPV: Negative predictive value

4 Youden index is calculated as follows: $(Se+Sp)-1$

5 Algorithms at 90%, 80% and 70% refer to the factors selected in 90%, 80% and 70% of the bootstrap
 6 samples for their construction

7

8

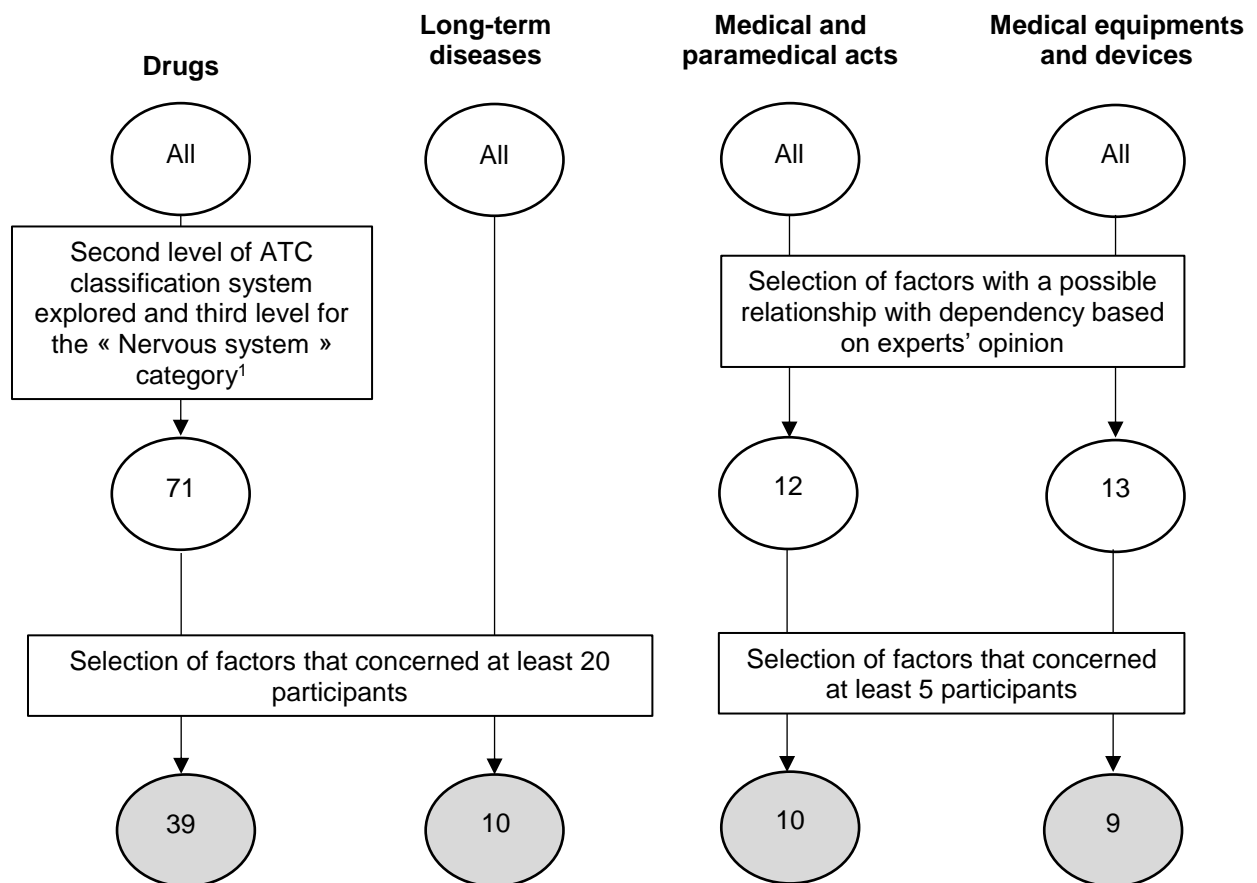


Figure 1. Selection of factors included in the analysis

The figure should be read as follows: among all medical and paramedical acts, 12 were selected by experts' opinion and, after exclusion of factors that concerned less than 5 participants, 10 were included in the LASSO analysis.

¹ In the ATC (Anatomical Therapeutic Chemical) classification drugs are divided into groups according to the organ or system on which they act or their therapeutic and chemical characteristics; drugs are coding on seven characters (letters and digits). The first level (first letter) defines the anatomical group among 14 different ones. The second level (first two digits) gives the main pharmacological or therapeutic subgroup and the third level (second letter) corresponds to chemical, pharmacological or therapeutic subgroups. In the analysis we used the second level but for drugs of the nervous system where the third level was considered due to their potential high relation with dependency.

Application of the algorithm to all persons

Algorithm = \sum (coefficients \times variables)

0.78 \times Age.....	75-<85 = 1
1.63 \times Age.....	$\geq 85 = 1$
0.24 \times Sex.....	woman = 1
1.30 \times Stroke.....	disease during the quarter or before = 1
1.03 \times Severe heart failure.....	disease during the quarter or before = 1
0.77 \times Coronary disease.....	disease during the quarter or before = 1
3.12 \times Alzheimer's disease and other dementia.....	disease during the quarter or before = 1
3.12 \times Medical bed	≥ 1 delivery the year before the quarter = 1
6.29 \times Patient lifter	≥ 1 delivery the year before the quarter = 1
1.28 \times Equipment for incontinence.....	≥ 1 delivery the year before the quarter = 1
2.90 \times Non daily nursing care.....	45 to 90 visits per quarter = 1
2.85 \times Daily nursing care.....	≥ 90 visits per quarter = 1
2.45 \times Ambulance.....	≥ 3 per quarter = 1
0.84 \times At-home GP' visits.....	≥ 3 per quarter = 1
2.51 \times Dopaminergic agents for Parkinson.....	≥ 2 deliveries per quarter = 1
0.96 \times Antidepressant drugs.....	≥ 2 deliveries per quarter = 1
0.81 \times Drugs used in diabetes.....	≥ 2 deliveries per quarter = 1
- 0.77 \times Lipid modifying agents.....	≥ 2 deliveries per quarter = 1

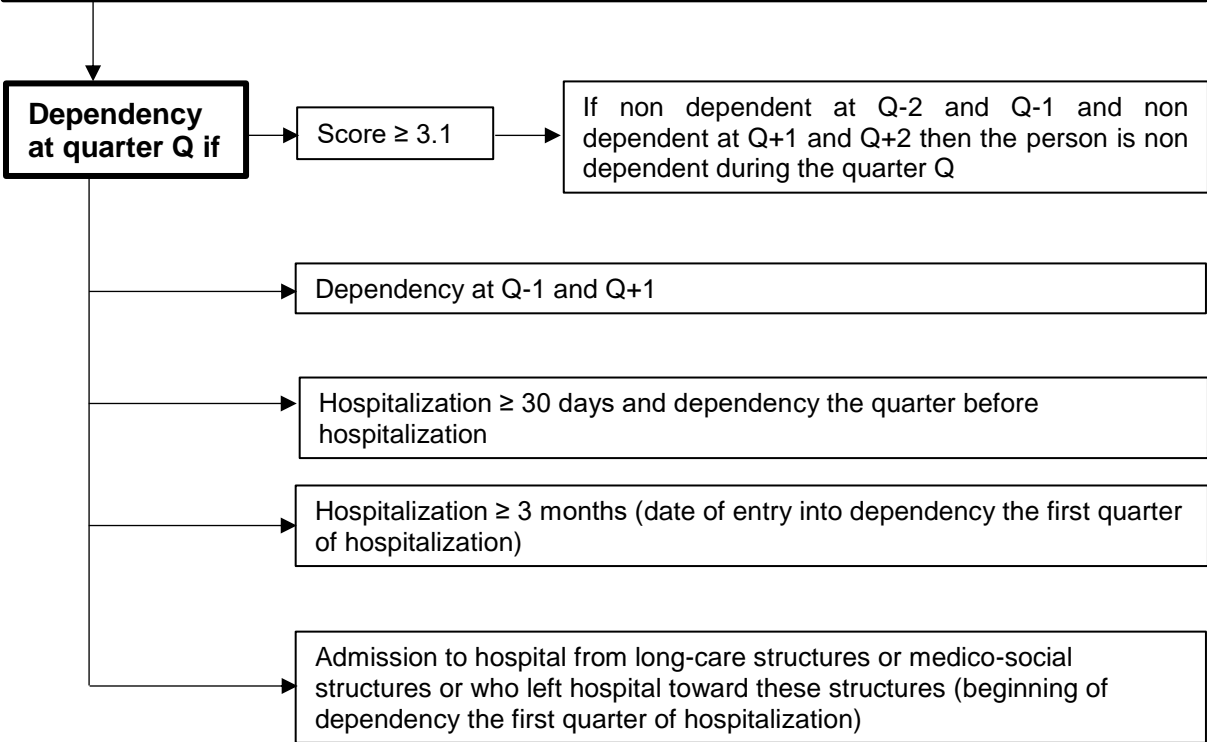


Figure 2. Algorithm to identify quarters of ADL-dependency

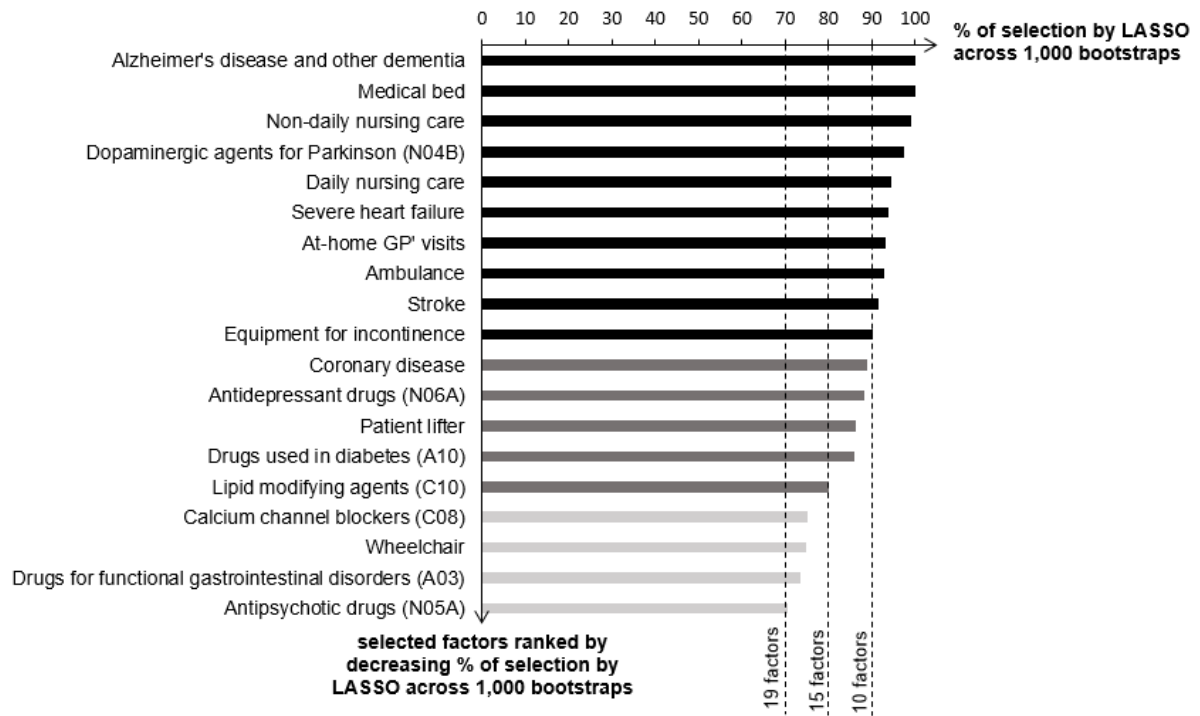


Figure 3. Factors selected by LASSO bootstraps

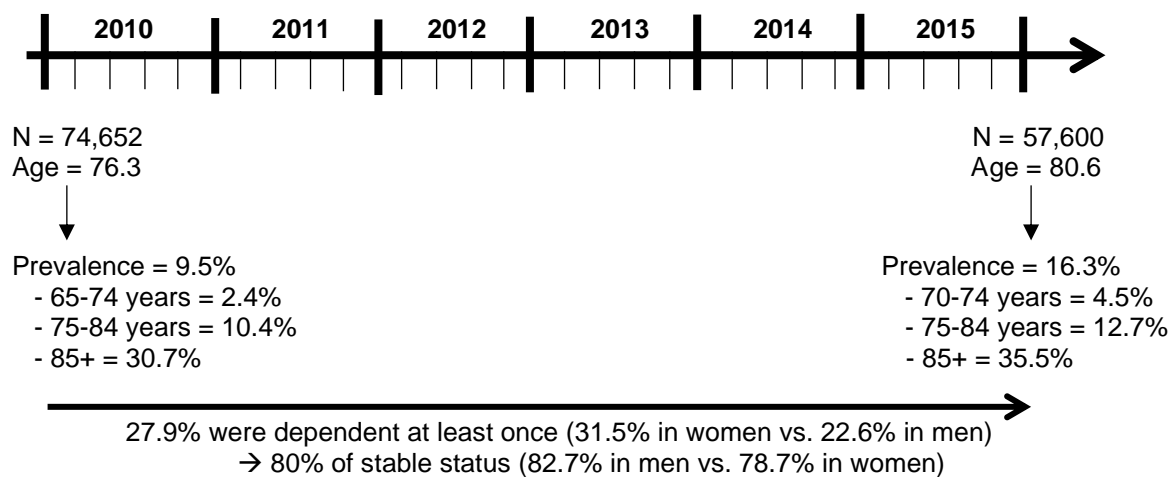


Figure 4. Prevalence of ADL-dependency among the EGB patients in 2010 first quarter and 2015 last quarter